フォルマント周波数軌跡を潜在パラメータとした音声スペクトル生成過程の確率モデル*

☆吉里幸太 [1]，北条伸克 [1]，亀岡弘和 [1,2]，齋藤大輔 [1]，嵯峨山茂樹 [1]
([1] 東大院・情報理工, [2]NTT CS 研)

1 はじめに

人間らしい音声合成を実現するにあたっては，音声のダイナミクスに現れる非言語情報やパラ言語情報を詳細にモデリングすることが重要である．例えば，音素特徴量の動的特徴には話者の個人性が現れていることが知られており [1]，音声認識，話者認識，音声合成などにおいて重要な特徴量の一つとして扱われている [1, 2, 3, 4]．一方，音声の基本周波数軌跡には韻律的な特徴が現われていることが知られており [5]，韻律解析や音声合成において基本周波数の動的特徴や動的モデルが重要な役割を果たしてきた [3, 5]．

また，近年利用可能な音声データベースが増加してきたことに伴って，統計的手法を用いた音声処理に関して盛んに研究が行われている．とくに音声の生成過程を確率モデルとして表現するアプローチは強力で，パラメータ推定に様々な統計的手法を活用できたり，先験的情報をパラメータの事前分布としてモデル内に組み込むことができたり，パラメータの統計的な振る舞いや傾向を学習することができたりといった様々な利点がある．

我々は，それらの重要性に注目して音声ダイナミクスの確率モデル化に取り組んでおり，その一環としてこれまでイントネーションの確率モデル化の検討を進めてきた [6, 7]．現在，我々の研究室では複合ウェーブレットモデル（Composite Wavelet Model; CWM）と呼ぶスペクトル包絡モデルに基づく新しい音声合成方式を検討中である [8, 9]．CWM は LSP のようにフォルマント周波数に対応していると解釈できるパラメータを有しており，当該音声合成方式へ将来的に組み込んでいくことを見据え，フォルマント周波数軌跡のダイナミクスの確率モデル化を行おうというのが本研究の動機である．

本稿ではまず，フォルマント周波数の時間軌跡を潜在パラメータとしてもつ音声スペクトル生成過程の確率モデルの定式化を行い，そのパラメータ推定アルゴリズムについて述べる．そして，実音声のスペクトル包絡からフォルマント周波数の時間軌跡を推定する実験を通して，提案手法の有効性を確認する．

2 音声スペクトル生成過程の確率モデル化

2.1 フォルマント軌跡の生成過程に関する仮説

フォルマント（本稿ではスペクトル包絡のピークと定義する）は，音声を特徴づける極めて重要な要素である．声帯振動が共振することによって生じるフォルマントの周波数軌跡には声道の運動に伴う何らかの物理的な制約が付随すると考えられるが，本研究では，フォルマント周波数軌跡が藤崎の F_0 パターン生成過程モデル [5] と同様のメカニズムによって生じると仮定する．具体的には，フォルマント周波数の対数の時間軌跡を，Fig. 1 に示すように音素区間ごとに一定値をとる階段状の指令関数（以後，音素指令関数）にインパルス応答

$$G(t) = \begin{cases} \alpha^2 t e^{-\alpha t} & (t \geq 0) \\ 0 & (t < 0) \end{cases} \qquad (1)$$

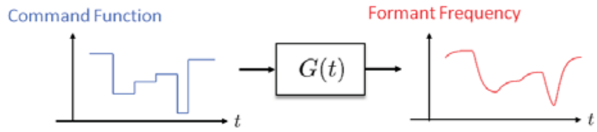

Fig. 1 線形系によるフォルマント周波数軌跡の生成過程

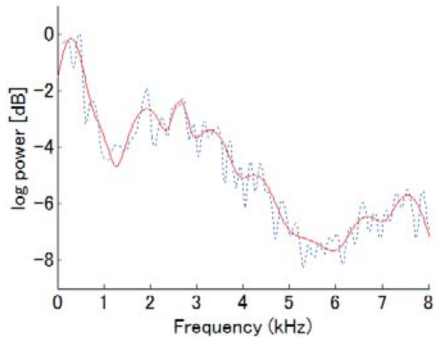

Fig. 2 観測信号 (音素/e/) のスペクトル包絡（点線）と GMM によるその近似（実線）の例（混合数 10）

が畳み込まれ（α は固有角周波数），二次線形系の出力として生じたものと考える．なお，二次線形系の仮定が置かれた他の音声生成過程モデルの例として，音素認識を目的とした調音運動の動的モデルが提案されている [10]．

実音声から直接観測できるのはスペクトル包絡であり，フォルマント周波数は陽には観測されない．そこで以下では，フォルマント周波数軌跡からどのようなプロセスを通して実際にスペクトル包絡が生成されるかについて議論し，そのプロセスを確率モデルとして定式化する．

2.2 複合ウェーブレットモデル [11, 12] の導入

スペクトル包絡における各フォルマントをガウス分布関数で近似的に表現できるとすると，スペクトル包絡全体をガウス分布関数の重ね合わせ，すなわち混合ガウス分布関数モデル（Gaussian Mixture Model; GMM）で表現することができる．スペクトル包絡の GMM による近似の例を Fig. 2 に示す．このようなスペクトル包絡の表現を，複合ウェーブレットモデル（Composite Wavelet Model; CWM）[11, 12] と呼ぶ．CWM におけるスペクトル包絡モデル $F_{\omega,t}$ は

$$\phi_{\omega,t} = \sum_{k=1}^{K} \psi_{k,\omega,t} \qquad (2)$$

$$\psi_{k,\omega,t} = \frac{w_{k,t}}{\sqrt{2\pi}\sigma_{k,t}} \exp\left(-\frac{(\omega-\mu_{k,t})^2}{2\sigma_{k,t}^2}\right) \qquad (3)$$

で与えられる．ただし，ω, t は周波数と時刻のインデックス，k はガウス関数のインデックスであり，K

---

* Probabilistic model of speech spectral sequences involving formant frequency contours as latent variables. by YOSHIZATO Kota, HOJO Nobukatsu, KAMEOKA Hirokazu, SAITO Daisuke, SAGAYAMA Shigeki (The University of Tokyo)

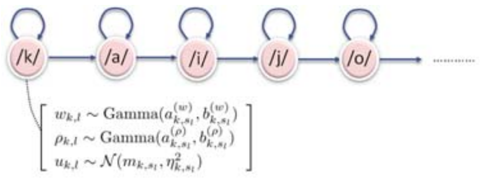

Fig. 3　提案 HMM の構成

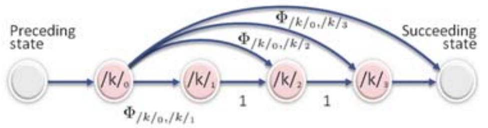

Fig. 4　状態/k/の 4 つの小状態への分割

は GMM の混合数である．また，$\mu_{k,t}, \sigma^2_{k,t}, w_{k,t}$ は，それぞれガウス分布関数を統計分布と見なした際の平均・分散・重みに対応し，フォルマント周波数・フォルマントピークの鋭さ・フォルマントの強度に対応するパラメータである．

2.3　確率モデルの定式化

本節では，以上の準備をもとに，フォルマント周波数の時間軌跡を潜在パラメータとしてもつ音声スペクトル生成過程の確率モデルを定式化する．

[6] のアイディアと同様，階段状の関数が隠れマルコフモデル（Hidden Markov Model; HMM）により表現できることに着目すると，音素境界により段が切り替わる階段状の音素指令関数を，Fig. 3 のような音素に対応した状態からなる HMM により確率モデル化することができる．これは，将来的に提案モデルを HMM 音声合成の枠組みに組み込んでいくことを考慮しても好都合である．すなわち，この HMM は，各離散時刻 l において CWM パラメータである $w_{k,l}$，$\sigma_{k,l}$（便宜的に以後 $\sigma_{k,l}$ の逆数を $\rho_{k,l}$ と置き，$\sigma_{k,l}$ の代わりに $\rho_{k,l}$ をパラメータと見なす。），および，音素指令関数 $u_{k,l}$ を出力する確率的ジェネレータと見なせる．

加えて，自己遷移の持続長をパラメータ化するために，隠れセミマルコフモデル (Hidden Semi-Markov Model; HSMM) [13] を導入する．HSMM は，各状態を十分大きな数の小状態に分割することと等価である．ここで，分割後の各小状態はすべて同じ出力分布を持つ．Fig. 4 に状態/k/を分割した例を示した．このような分割により，ある状態にある離散時間だけ留まる確率を個別にパラメータ化することが可能になる．以上より，提案 HMM の構成は以下となる．

出力値系列: $\{w_{k,l}, \rho_{k,l}, u_{k,l}\}_{k,l}$
状態集合: $\{/\mathrm{a}/_i, /\mathrm{k}/_i, /\mathrm{o}/_i, \ldots\}_i$
状態系列: $\boldsymbol{s} = \{s_l\}_l$
状態出力分布:

$$
\begin{aligned}
P(w_{k,l}|s_l) &= \mathrm{Gamma}(w_{k,l}; a^{(w)}_{k,s_l}, b^{(w)}_{k,s_l})\\
P(\rho_{k,l}|s_l) &= \mathrm{Gamma}(\rho_{k,l}; a^{(\rho)}_{k,s_l}, b^{(\rho)}_{k,s_l})\\
P(u_{k,l}|s_l) &= \mathcal{N}(u_{k,l}; m_{k,s_l}, \eta^2_{k,s_l})
\end{aligned}
$$

状態遷移確率: $\Phi_{i',i} = P(s_l = i | s_{l-1} = i')$
初期状態確率: $\Phi_i = P(s_1 = i)$

ただし，$\mathrm{Gamma}(x; a, b)$ はガンマ分布

$$
\mathrm{Gamma}(x; a, b) = x^{a-1}\frac{\exp(-x/b)}{\Gamma(a)\ b^a} \qquad (4)
$$

である．

2.1 節で議論したように，提案モデルでは指令関数 $u_{k,l}$ に二次線形系のインパルス応答が畳み込まれてガウス分布関数の平均値の軌跡 $\mu_{k,l}$ が生じるとする．具体的には

$$
P(\mu_{k,l}|\{u_{k,l}\}_l, s_l) = \mathcal{LN}(G_{k,l} \overset{l}{*} u_{k,l}, \nu^2_{k,s_l}), \qquad (5)
$$

と書けるとする．ここで $\mathcal{LN}(x; \mu, \sigma^2)$ は対数正規分布であり，$\log x$ が正規分布 $\mathcal{N}(x; \mu, \sigma^2)$ に従うことと等価である．また $G_{k,l}$ は式 (1) の $G(t)$ の離散時間表現であり（固有周波数は α_k とおく），$\overset{l}{*}$ は離散時刻に関する畳み込みを表す．以降，パラメータをまとめて $\boldsymbol{\rho} = \{\rho_{k,l}\}_{k,l}, \boldsymbol{w} = \{w_{k,l}\}_{k,l}, \boldsymbol{u} = \{u_{k,l}\}_{k,l}, \boldsymbol{\mu} = \{\mu_{k,l}\}_{k,l}$ と書く．

すべての CWM パラメータと状態系列 $\boldsymbol{s}$ が与えられたときに観測スペクトル包絡 $y_{\omega,l}$ が生じる確率を

$$
P(y_{\omega,l}|\boldsymbol{\rho}, \boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{s}) = \mathrm{Poisson}(y_{\omega,l}; \phi_{\omega,l}) \qquad (6)
$$

とする．ここで $\mathrm{Poisson}(x; \lambda)$ はポアソン分布である．なお，この仮定の下での λ の最尤推定問題は，スペクトル間の近さを測る尺度の一つとして近年音響信号処理分野で多用される I ダイバージェンスと呼ぶ歪み尺度を規準とした x と λ の最適フィッティング問題と等価となることが知られている [14]。

以上の提案モデルの推定すべきパラメータをまとめて $\Theta = \{\boldsymbol{\rho}, \boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{u}, \boldsymbol{s}, \boldsymbol{\theta}\}$（ただし，$\boldsymbol{\theta} = \{b^{(w)}_{k,i}, b^{(\rho)}_{k,i}, m_{k,i}\}_{k,i}$）とし，次章では，観測スペクトル系列 $Y = \{y_{\omega,l}\}_{\omega,l}$ が与えられた下での事後確率 $P(\Theta|Y)$ を最大化するパラメータ推定アルゴリズムについて述べる．

3　パラメータ推定アルゴリズム

簡単のため，本稿ではパラメータ $\{a^{(w)}_{k,i}, a^{(\rho)}_{k,i}, \eta^2_{k,i}, \nu^2_{k,i}\}_{k,i}, \{\alpha_k\}_k$ をすべて定数とする．また，$\boldsymbol{\theta}$ の事前分布は一様分布とする．

$P(\Theta|\boldsymbol{y})$ を最大化する Θ を解析的に求めることは難しいが，以下に述べるように補助関数法に基づき局所最適化アルゴリズムを導くことができる．

$\log P(\Theta|\boldsymbol{y})$ は，

$$
\log P(\Theta|\boldsymbol{y}) \overset{c}{=} \log P(\boldsymbol{y}|\Theta)P(\Theta) \qquad (7)
$$

$$
\begin{aligned}
\log P(\Theta) \overset{c}{=} &\log P(\boldsymbol{s})P(\boldsymbol{\rho}|\boldsymbol{s}, \boldsymbol{\theta})P(\boldsymbol{w}|\boldsymbol{s}, \boldsymbol{\theta})\\
&P(\boldsymbol{u}|\boldsymbol{s}, \boldsymbol{\theta})P(\boldsymbol{\mu}|\boldsymbol{s}, \boldsymbol{u}, \boldsymbol{\theta}) \qquad (8)
\end{aligned}
$$

と表される．ただし，$\overset{c}{=}$ は定数部分を除いた場合の等号を意味する．先述のように，$-\log P(\boldsymbol{y}|\Theta)$ は定数項を除けば観測スペクトル包絡 $y_{\omega,l}$ とスペクトル包絡モデル $\phi_{\omega,l}$ との間の I ダイバージェンスと等しく [14]，さらに，

$$
\begin{aligned}
\sum_{\omega} \psi_{k,\omega,l} &\simeq \int_{-\infty}^{\infty} \frac{w_{k,l}}{\sqrt{2\pi}\sigma_{k,l}} \exp\left(-\frac{(\omega - \mu_{k,l})^2}{2\sigma^2_{k,l}}\right) \mathrm{d}\omega\\
&= w_{k,l} \qquad (9)
\end{aligned}
$$

となることを用いれば,

$$-\log P(\boldsymbol{y}|\Theta) \stackrel{c}{=} \sum_{\omega,l}\left(y_{\omega,l}\log\frac{y_{\omega,l}}{\phi_{\omega,l}} - y_{\omega,l} + \phi_{\omega,l}\right)$$
$$\stackrel{c}{=} \sum_{\omega,l}(\phi_{\omega,l} - y_{\omega,l}\log\phi_{\omega,l})$$
$$\simeq \sum_{k,l} w_{k,l} - \sum_{\omega,l} y_{\omega,l}\log\phi_{\omega,l} \tag{10}$$

が言える．この式の $-y_{\omega,l}\log\phi_{\omega,l}$ の項に対し，負の対数関数の凸性を利用し、Jensen の不等式を用いることで

$$-y_{\omega,l}\log\phi_{\omega,l} \leq -y_{\omega,l}\sum_{k}\gamma_{k,\omega,l}\log\frac{\psi_{k,\omega,l}}{\gamma_{k,\omega,l}} \tag{11}$$

のように上界関数を設計することができる．また，$-\log P(\boldsymbol{\mu}|\boldsymbol{s},\boldsymbol{u})$ の $\left(\sum_\tau G_{k,l-\tau}u_{k,\tau}\right)^2$ の項に対し，二次関数の凸性を利用し，同様に Jensen の不等式を用いることで

$$\left(\sum_{\tau} G_{k,l-\tau}u_{k,\tau}\right)^2 \leq \sum_{\tau}\frac{(G_{k,l-\tau}u_{k,\tau})^2}{\lambda_{\tau,k,l}} \tag{12}$$

のように上界関数を設計することができる [15]．さらに $-\log P(\boldsymbol{\mu}|\boldsymbol{s},\boldsymbol{u})$ の中の $(\log\mu_{k,l})^2$ の項に関しては,

$$(\log\mu_{k,l})^2 \leq \frac{1}{\mu_{k,l}} + \left(\frac{2\log\xi_{k,l}}{\xi_{k,l}} + \frac{1}{\xi_{k,l}^2}\right)\mu_{k,l}$$
$$+|\log\xi_{k,l}|^2 - 2\log\xi_{k,l} - \frac{2}{\xi_{k,l}} \tag{13}$$

のように上界関数を設計することができる [16]．ここで $\gamma_{k,\omega,l}, \lambda_{\tau,k,l}, \xi_{k,l}$ は補助変数である．

式 (11)〜(13) より，$-\log p(\Theta|Y)$ の上界関数を設計することができ，これを補助関数として補助関数法を適用することができる．まず，補助変数の更新式は，上述の不等式の等号成立条件，すなわち，

$$\lambda_{\tau,k,l} = \frac{G_{k,l-\tau}u_{k,\tau}}{\sum_{\tau'} G_{k,l-\tau'}u_{k,\tau'}} \tag{14}$$
$$\gamma_{k,\omega,l} = \frac{\psi_{k,\omega,l}}{\phi_{\omega,l}} \tag{15}$$
$$\xi_{k,l} = \mu_{k,l} \tag{16}$$

で与えられる．モデルパラメータ Θ の更新式は，補助関数を最小化する解，すなわち偏微分が 0 となる解となる．$\boldsymbol{\mu}$ 以外のパラメータの更新式は,

$$m_{k,i} = \frac{\sum_{l\in\mathcal{T}_i}\frac{u_{k,l}}{\eta_{k,s_l}^2}}{\sum_{l\in\mathcal{T}_i}\frac{1}{\eta_{k,s_l}^2}} \tag{17}$$

$$b_{k,i}^{(\rho)} = \frac{\sum_{l\in\mathcal{T}_i}\rho_{k,l}}{\sum_{l\in\mathcal{T}_i}a_{k,s_l}^{(\rho)}},\quad b_{k,i}^{(w)} = \frac{\sum_{l\in\mathcal{T}_i}w_{k,l}}{\sum_{l\in\mathcal{T}_i}a_{k,s_l}^{(w)}} \tag{18}$$

$$u_{k,l} = \frac{\frac{m_{k,s_l}}{\eta_{k,s_l}^2} + \sum_{\tau\geq l}\frac{G_{k,\tau-l}\log\mu_{k,\tau}}{\nu_{k,s_\tau}^2}}{\frac{1}{\eta_{k,s_l}^2} + \sum_{\tau\geq l}\frac{G_{k,\tau-l}^2}{\nu_{k,s_\tau}^2\lambda_{l,k,\tau}}} \tag{19}$$

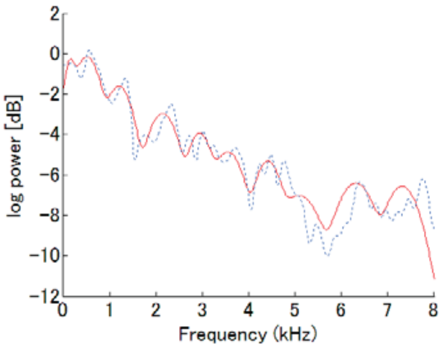

Fig. 6　No.436 の音素/a/のスペクトル包絡（点線）の GMM 近似（実線）

$$\rho_{k,l} = \frac{2(a_{k,s_l}^{(\rho)} - 1) + \sum_\omega y_{\omega,l}\gamma_{k,\omega,l}}{\frac{2}{b_{k,s_l}^{(\rho)}} + \sum_\omega y_{\omega,l}\gamma_{k,\omega,l}(\omega-\mu_{k,l})^2} \tag{20}$$

$$w_{k,l} = \frac{a_{k,s_l}^{(w)} - 1 + \sum_\omega y_{\omega,l}\gamma_{k,\omega,l}}{\frac{1}{b_{k,s_l}^{(w)}} + 1} \tag{21}$$

となる．ここで $\mathcal{T}_i = \{l|s_l = i\}$ である．$\boldsymbol{\mu}$ については,

$$p_3\mu_{k,l}^3 + p_2\mu_{k,l}^2 + p_1\mu_{k,l} + p_0 = 0 \tag{22}$$
$$p_3 = \sum_{\omega} y_{\omega,l}\gamma_{k,\omega,l}\rho_{k,l} \tag{23}$$
$$p_2 = \frac{2\xi_{k,l}\log\xi_{k,l} + 1}{2\nu_{k,s_l}^2\xi_{k,l}^2} - \sum_{\omega} y_{\omega,l}\gamma_{k,\omega,l}\rho_{k,l}\omega \tag{24}$$
$$p_1 = 1 - \frac{1}{\nu_{k,s_l}^2}\sum_{\tau\leq t} G_{k,l-\tau}u_{k,\tau} \tag{25}$$
$$p_0 = -\frac{1}{2\nu_{k,s_l}^2} \tag{26}$$

とおいて，式 (22) の正の解のうち $-\log P(\Theta|\boldsymbol{y})$ を最も小さくする $\mu_{k,l}$ を選べばよい．

以上の更新則を十分な回数反復することで，$P(\Theta|\boldsymbol{y})$ を局所最大化するパラメータ Θ を推定することができる．

4　実験

提案モデルが実音声のフォルマント周波数の時間軌跡をよく表現できていることを確認するために実験を行った．実験は大きく分けて学習フェイズと推定フェイズの 2 段階からなる．学習フェイズでは，ATR 日本語音声データベースの B セット [17] から男性話者 1 人を選択し，No.1〜No.400 までの 400 文を対象として，音素ごとに定まるパラメータ $\boldsymbol{\theta}$ の学習を行った．続く推定フェイズでは，学習に使っていない発話文を対象に CWM パラメータの推定を行った．ここで $\boldsymbol{\theta}$ は学習フェイズでの推定値を用いて定数とみなした．なお，本実験においてスペクトル包絡の抽出には STRAIGHT[18] を用い，また音素ラベルのデータを与えることで状態系列 $\boldsymbol{s}$ は定数とした．

本実験では，GMM の混合数は 10，パラメータ推定アルゴリズムの反復回数は 10，$\alpha_k = 50$ とし，その他の CWM パラメータの初期値は [8] の Chain を導入しない推定アルゴリズムを用いて決定した．

No.436 の「鉛筆だと，力を入れて書くので，スピードがにぶるのである」の読み上げ音声を対象に推定を行った結果を Fig. 5 と Fig. 6 に示した．Fig. 5 は,

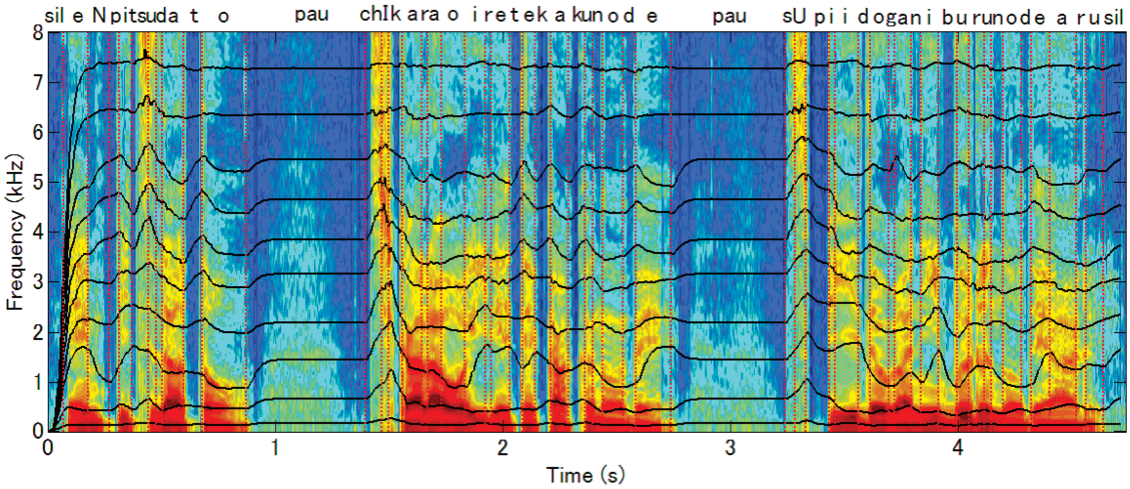

Fig. 5　No.436 のスペクトル包絡系列 (カラーマップ表示) と提案法による推定フォルマント周波数軌跡 (実線) (点線は音素境界)

No.436 のスペクトル包絡に提案手法で推定したフォルマント周波数の時間軌跡を重ねて描いた図である．また Fig. 6 は，同じく No.436 の音素/a/のスペクトル包絡と推定パラメータによる GMM 近似を示した図である．これらの結果から，推定したスペクトルピークが実音声のフォルマント周波数軌跡をうまく表現できていることが確認できる．

5　おわりに

本研究では，フォルマント周波数の時間軌跡を潜在パラメータとしてもつ音声スペクトル生成過程の確率モデルの定式化を行い，そのパラメータ推定アルゴリズムを導出した．そして実験を通して，提案手法が実音声のスペクトルピークをよく表現できることを確認した．

今後は，フォルマント周波数軌跡を二次線形系の出力とみなしたとき，固有角周波数 α に個人性が表れるのではないかといった，より詳細なモデルの検討を行なっていきたい．また，提案モデルを基本周波数パターンの生成過程の確率モデル [7] と統合することで，より高品質なテキスト音声合成を実現する研究を進めていくことを考えている．

参考文献

[1] 嵯峨山, 板倉, “音声の動的尺度に含まれる個人性情報,” 音講論 (春), No. 3-2-7, pp. 589–590, 1979.

[2] 古井, 齋藤, “音声スペクトルの動的特徴を用いた話者認識,” NTT 電気通信研究所研究実用化報告, Vol. 29, No. 7, pp. 1263–1276, 1980.

[3] 全他, “静的・動的特徴の明示的な関係によりHMM から導出されるトラジェクトリモデル,” 信学技報, SP2003-122, pp. 55–60, 2003.

[4] S. Furui, “Speaker-independent isolated word recognition using dynamic features of speech spectrum,” *IEEE Trans. Acoust., Speech, Signal Process.*, Vol. 34, No. 1, pp. 52–59, 1986.

[5] H. Fujisaki, In *Vocal Physiology: Voice Production, Mechanisms and Functions*, (O. Fujimura, ed.) Raven Press, pp. 347–355, 1988.

[6] 亀岡他, “音声 F_0 パターン生成過程の確率モデル,” 音講論 (秋), No. 1-1-3, pp. 207–210, 2010.

[7] 吉里他, “F_0 パターン生成過程の確率モデルによる藤崎モデルパラメータの推定,” 情処研報, Vol. 2012-SLP-92, No. 9, pp. 1–6, 2012.

[8] 北条他, “複合ウェーブレットモデル分析合成系に基づく HMM 音声合成,” 音講論 (秋), No. 2-2-7, pp. 287–290, 2012.

[9] 北条他, “複合ウェーブレットモデルと隠れマルコフモデルの統合モデルによるテキスト音声合成,” 音講論 (春), to appear, 2013.

[10] 誉田, “調音モデルにもとづく音声の特徴抽出に関する研究,” 博士論文, 早稲田大学, 1977.

[11] 槐他, “複合ウェーブレットモデルに基づく音声の分析合成,” 信学技報, Vol. 105, No. 372, pp. 1–6, 2005.

[12] P. Zolfaghari, T. Robinson, “Formant Analysis using Mixtures of Gaussians,” in *Proc. ICSLP’96*, Vol. 2, pp. 1229–1232, 1996.

[13] J. D. Ferguson, “Variable duration models for speech,” *Symposium on the Application of Hidden Markov Models to Text and Speech*, pp. 143–179, 1980.

[14] H. Kameoka, “Statistical Approach to Multipitch Analysis,” Ph.D. Thesis, The University of Tokyo, 2007.

[15] 亀岡他, “音声のスパース性と非負制約つき畳み込みモデルに基づくパワースペクトル領域残響除去,” 音講論 (秋), No. 3-8-10, pp. 705–708, 2008.

[16] H. Kameoka *et al.*, “Constrained and regularized variants of non-negative matrix factorization incorporating music-specific constraints,” in *Proc. ICASSP’12*, pp. 5365–5368, 2012.

[17] A. Kurematsu *et al.*, “ATR japanese speech database as a tool of speech recognition and synthesis,” *Speech Communication*, Vol. 27, pp. 187–207, 1999.

[18] H. Kawahara *et al.*, “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0extraction,” *Speech Communication*, Vol. 27, No. 3-4, pp. 187–207, 1999.